BELLCOMM, INC.
1100 SEVENTEENTH STREET, N.W. WASHINGTON, D.C. 20036

#### COVER SHEET FOR TECHNICAL MEMORANDUM

TITLE- On The Reliability Of Computer Operations

TM- 68-1033-5

DATE- September 13, 1968

FILING CASE NO(S)- 101

AUTHOR(S)- G. L. Yang

FILING SUBJECT(S)- Computed Do THE RELIABILITY OF (ASSIGNED BY AI (NASA-CR-73518) ON THE RELIABILITY OF COMPUTER OPERATIONS (Bellcomm, Inc.) 24 p

N79-73136

Unclas 00/60 11307

ABSTE

(NASA CR OR TMX OR AD NUMBER) (CATEGORY)

The usual assumptions postulated in a reliability study of a machine are the constant failure rate and constant repair rate (of the machine). The study of Bellcomm's computer UNIVAC 1108 indicates the fact that the hardware failures of the computer are time-dependent in occurrence. More precisely, the probability of a failure to occur varies with the time of day. This motivates the use of a non-stationary stochastic process to describe the reliability of the computer; its probability distribution is governed by the time-dependent failure rate and the constant repair rate. problem has been studied under two different situations: firstly, we assume that the failure rate is time-dependent but that the repair time is negligible, the model is then used to fit the Bellcomm computer hardware failure data over a period of three months, they are October, November of 1967 and March of 1968; secondly, we assume that the failure rate is time-dependent and the repair time follows an exponential distribution. Because of inadequate repair time data, only the derivation of the model is discussed here.

The reliability study covers the determination of the probability distributions of the following:

- 1. Number of failures in a specified time interval.
- 2. The availability of the computer at a given time or in a given interval. By the availability, we mean the successful operation of the machine and the expected length of time that the computer will operate successfully in a given time interval. Also included in the study is the estimation of the parameters involved in the model.

#### DISTRIBUTION

## COMPLETE MEMORANDUM TO

#### COVER SHEET ONLY TO

#### CORRESPONDENCE FILES:

OFFICIAL FILE COPY
plus one white copy for each
additional case referenced

#### TECHNICAL LIBRARY (4)

Messrs. T. A. Keegan - NASA/MA-2

J. T. McClanahan - NASA/MOR

W. E. Miller, Jr. - NASA/MOG

J. P. Nolan - NASA/MOA

L. M. Robinson - NASA/TS

Messrs. L. C. Dunseith - MSC/FS

L. R. Hamilton - MSC/FS3

H. E. Leech - MSC/FS3

L. L. Linson - MSC/FS3

J. B. MacLeod - MSC/FS5

J. A. Miller - MSC/FS2

D. H. Owen, Jr. - MSC/FC

### Bellcomm

Messrs. G. R. Andersen

D. R. Anselmo

H. J. Bixhorn

H. B. Bosch

A. P. Boysen, Jr.

J. O. Cappellari

K. R. Carpenter

D. A. Chisholm

D. A. Corev

B. H. Crane

J. P. Downs

W. B. Gevarter

D. R. Hagner

P. L. Havenstein

W. G. Heffron

H. A. Helm

J. J. Hibbert

E. E. Hillyard

B. T. Howard

J. E. Johnson

S. L. Leyle

B. H. Liebowitz

H. S. London

J. L. Marshall

J. Z. Menard

A. C. Merritt

V. S. Mummert

J. M. Nervik

J. T. Raleigh

P. E. Reynolds

I. M. Ross

P. S. Schaenman

F. N. Schmidt

R. L. Selden

R. V. Sperry

J. W. Timko

R. L. Wagner

All Members of Department 1033

Department 1024 File

The preceding consideration of one computer can easily be generalized to the reliability study of N(>1) identical computers. The following points are also discussed:

- 1. The probability distribution of the number of computers operating successfully at a given time.
- 2. The optimal number of computers required to guarantee the system's availability.

SUBJECT: On The Reliability Of Computer Operations - Case 101

DATE: September 13, 1968

FROM: G. L. Yang

TM-68-1033-5

### TECHNICAL MEMORANDUM

#### 1. INTRODUCTION

Hardware failures of the computer are observed and recorded on a daily basis as they occur along the time axis. The experiment starts at time 0, say, and ends at time T ( $<\infty$ ). Let  $T_i$  be the time of the  $i^{th}$  failure. After each failure  $T_i$  there follows a period of down time during which the computer is being repaired. If  $\xi_i$  denotes the recovery time of the computer, then the repair time of the  $i^{th}$  failure is  $\xi_i - T_i = V_i$ . For some failures the repair time  $V_i$  is negligible, while for others it is lengthy. For clarity, we present our data on the following time axis

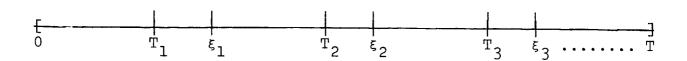


Figure 1

with 0 <  $T_1$  <  $\xi_1$  <  $T_2$  <  $\xi_2$  <  $T_3$  <  $\xi_3$  .... < T, the i<sup>th</sup> repair time being  $\xi_i$  -  $T_i$  for i = 1, 2, ... Let (a,b) denote an interval with a < b. If there are n failures in [0, T], then the available time of the computer is given by

$$\bigcup_{i=1}^{n} (\xi_{i-1}, T_i) \bigcup (\xi_n, T)$$

where  $\xi_0 = 0$  and  $(\xi_n, T) = \emptyset$  if  $\xi_n > T$ .

Consider one computer, let N(t) be the number of failures occurred "up to time t" given that we start to observe the computer operation at time  $t_0=0$ . We also introduce X(t) to describe the state of the computer at time t. X(t) is equal to 1 if the computer is operating at time t and is otherwise equal to zero. In the following sections, the processes N(t) and X(t) will be investigated under various assumptions.

## 2. RELIABILITY MODEL UNDER THE ASSUMPTION OF INSTANTANEOUS REPAIR TIME

The main concern of this section is to study the process N(t),  $t \in [0,T]$ . Let  $P_n(t) = P[N(t)=n]$  be the probability of having n failures in the time interval [0,t]. For a fixed t, N(t) assumes values 0, 1, 2, ... Regarding the instantaneous transition of the N(t) process, we assume that the conditional probability that  $N(t+\Delta)$  equals n given that N(t) equals n-1 at time t is given by

$$P[N(t+\Delta) = n/N(t)=n-1] = \lambda(t)\Delta + o(\Delta), \Delta \rightarrow 0$$

where  $\Delta \to 0$ , and  $\lambda(t)$  is positive and a continuous function in t. The possible transitions of the process N(t) in a small time interval (t, t+ $\Delta$ ) is either increasing the failure by one (say from n-l to n) or remains in state n-l through time  $\Delta$ . The probability of multiple failures occurring in time  $\Delta$  is of smaller order of  $\Delta$ . The function  $\lambda(t)$  may be interpreted as the failure occurrence rate at time t, which is time-dependent. With these preliminaries, we can derive easily the differential-difference equations for the probability  $P_n(t)$ .

$$\frac{dP_n(t)}{dt} = -\lambda(t)P_n(t) + \lambda(t)P_{n-1}(t) \quad \text{for } n \ge 1$$

$$\frac{dP_0(t)}{dt} = -\lambda(t)P_0(t) ,$$

with the initial condition  $P_0(0) = 1$ . To solve equation (1) for  $P_n(t)$ , we introduce the following probability generating function G(s,t).

(2) 
$$G(s,t) = \sum_{n=0}^{\infty} s^n P_n(t) \qquad \text{for } |s| \le 1$$

Multiplying equation (1) by  $s^n$  and summing over n, we reduce equation (1) to a partial differential equation.

(3) 
$$\frac{\partial G(s,t)}{\partial t} = -\lambda(t) G(s,t) + \lambda(t) s G(s,t)$$
 for  $|s| \le 1$ 

With the corresponding initial condition G(s,0) = 1. The solution of G(s,t) is immediate, it is given by

$$G(s,t) = \exp \{-(1-s) \int_0^t \lambda(z) dz\} \text{ for } |s| \le 1$$
(4)
$$= \exp \{-(1-s)a(t)\} \text{ , with } a(t) = \int_0^t \lambda(z)dz \text{ .}$$

Equation (4) is a generating function of a Poisson random variable. Hence we conclude that, for every fixed t, the probability distribution of N(t) is given by

(5) 
$$P_n(t) = P[N(t)=n] = \frac{e^{-a(t)}[a(t)]^n}{n!}$$
 for n=0, 1, 2, ...

The process N(t) is in general a non-stationary Poisson process with the parameter set [0,T],  $T<\infty$ . Note that if a(t)=at (a is a positive constant), then N(t) reduces to a time homogeneous Poisson process.

Our next problem is to estimate the parameter  $a(t) = \int_0^t \lambda(z)dz$ .

For convenience and clarity, we use our computer data to illustrate the estimation procedure. The numerical results will be given after the explanation of the estimation method.

Our computer is operated on a daily basis; it is shut down on Saturdays and Sundays. The time from 4 to 6 a.m. of every working day is the maintenance and check up period and no observations are taken. Since occasionally the maintenance period extends beyond 6 a.m., we take 8 a.m. to be the starting time of our experiment. We thus have the observation period [0,T] to be [8 a.m., 4 a.m.]

Let I be the total number of days we have observed, and let  $N_{i}(t)$  be the number of failures up to time t on the  $i^{th}$  day for t  $\epsilon$  [0,t]. The observed  $N_{i}(t)$  represents a sample path of the stochastic process N(t) defined in equation (5). By our construction, the processes  $N_{i}(t)$  are independently distributed for  $i=1,2,\ldots,I$ .

For any fixed t, t  $\epsilon$  [0,T], the probability distribution of N  $_{\!\!\!1}(t)$  is given by

$$P[N_{i}(t)=n_{i}] = \frac{e^{-a(t)} a^{n_{i}(t)}}{(n_{i})!}$$
,

the same as that of equation (5) and the joint probability distribution of  $N_i(t)$ ,  $i=1, \ldots, I$  is the product

$$P[N_{i}(t) = n_{i}(t), i=1,...,I] = \prod_{i=1}^{I} P[N_{i}(t)=n_{i}(t)]$$

$$e^{-Ia(t)}[a(t)]^{i=1} n_{i}(t) \prod_{i=1}^{I} n_{i}(t) !$$

Thus, we have the logarithm of likelihood function is equal to

$$-Ia(t) + \sum_{i=1}^{I} n_i(t) \log a(t)$$

The maximum likelihood estimate  $\hat{a}(t)$  of a(t), for a fixed t, can be easily obtained.

$$\hat{a}(t) = \frac{\sum_{i=1}^{I} N_{i}(t)}{I} = \overline{N}(t) \quad t \in [0,T]$$

Since  $N_i(t)$  are independently and identically distributed, with  $EN_i(t) = a(t) < \infty$  and variance of  $N_i(t) = a(t) < \infty$ . By the law of large numbers, for every fixed t,

$$\hat{a}(t) = \frac{\sum_{i=1}^{I} N_{i}(t)}{I}$$

converges to a(t) a.s. as I tends to infinity. Thus we have  $\hat{a}(t)$  a consistent and obviously an unbiased estimator of a(t).

# 3. THE NUMERICAL RESULTS OF UNIVAC 1108 DATA COMPUTED UNDER THE MODEL OF SECTION 2

The hardware failure data were collected by Department 1032 during the months of October and November of 1967 and March of 1968. The observations were taken on a daily basis from 8 a.m. to 4 a.m. of the next day for a period of 20 hours. The failure data are summarized in frequency tables (I), (II), and (IV) for the three months separately.

In Table (I), we divided the time axis into hourly intervals using the beginning of each hour as diving points. In the first part of Table (I), we have plotted the number of failures N. (see definition in section 1) in each hour over the whole month (October). However, in the original raw data, we noticed that occasionally, the same failure source was recorded more than

once due to the fact that the failure was not properly repaired at the first time that the observer located the failure. Eliminating those "duplicated" failures, we plotted only "cluster centers" of these failures. This is given in the lower part of Table (I). In both parts of Table (I) we see that the failure frequencies vary with time, there are approximately three frequency peaks which occurred in the sub-intervals from 9 to 10 a.m., 13 to 14 p.m. and 16 to 17 p.m. The failure rate is obviously non-linear.

In order to check the difference in the pattern of failure frequency if we use another dividing system to obtain the hourly intervals, we chose 25 minutes after each hour as the dividing points of the sub-intervals. As shown in Table II (of October data), the non-linear failure frequencies persist, however the peaks are different. This indicates that the failure frequency pattern is sensitive to the different choice of dividing points of sub-intervals (of the same length). However, this causes no difficulty in the analysis, since we employ a continuous-time stochastic process.

From now on, for consistency, the dividing points of sub-intervals are chosen to be 8:25, 9:25, 10:25, ..., and so on. The November (Table III) and March data (Table IV) showed a similar pattern of failure frequencies. We observe the highest peak near noon time and two local peaks in the early morning and late afternoon respectively. The total number of failures decreases gradually from 72 cases in October to 49 cases in November and further to 42 cases in March. This is due to the gradual settlement of the newly installed computer.

Whatever may be the peaks, what we like to emphasize is the non-linear feature of the failure data. Using the model developed in section 2, we have

$$P[N(t)=n] = \frac{a^{n}(t) e^{-a(t)}}{n!}$$
 for n=0, 1, ...

with a(t) denoting the expected number of failures in

$$[0,t] = \int_0^t \lambda(z) dz.$$

As a comparison, we present the usual analysis by using a constant failure rate  $\lambda$ . Corresponding to this case, we have

(6) 
$$P[N(t)=n] = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$
 for n=0,1,2, ...

and the expected number of failures in (0,t) is  $\lambda t$ . The estimate  $\hat{a}(t)$  of a(t) is plotted in Table (X) on the basis of the estimation method derived in section 2. The broken line (ii) is the estimate based on the linear model (equation (6)), where the estimate  $\hat{\lambda}$  of

$$\lambda$$
 is calculated from the formula  $\hat{\lambda} = \frac{\sum\limits_{i=1}^{I} N_i(T)}{IT}$  .

Comparing the two estimates  $\hat{a}(t)$  and  $\hat{\lambda}t$  in Table (X) by "eye" inspection, we see that they give quite different results. The non-linear variation revealed in  $\hat{a}(t)$  is rather obvious. Of course, we could have made the comparison more rigorous by performing a test of hypothesis of constant failure rate  $\lambda$  against a time dependent one  $\lambda(t)$ . This might be done in the future, however, it does not seem necessary at the present (see Conclusion).

From Table (X) we can read off  $\hat{a}(t)$  and calculate the probability P[N(t)=n] accordingly. As an example, we given in Table (XI) the probabilities corresponding to  $\hat{a}(t)=1.4$  where t is equal to 15:10 p.m. The probabilities may be read from Biometrika Table [1].

Even an arbitrary straight line such as (i) through the origin to the  $\hat{a}(t)$  curve, would be a better fit than line (ii) at least up to time t=22:25. For line (i), at t=15:10 p.m., we have  $\lambda t=1$ , which differs from  $\hat{a}(t)=1.4$  by .4. The difference of the corresponding probabilities P[N(t)=n] is given in Table (XI). It shows that the linear model results could deviate in probability (see column (3)) from the observed - estimated data by 50% (see n=0).

## 4. RELIABILITY STUDY OF ONE COMPUTER WITH TIME-DEPENDENT FAILURE RATES AND CONSTANT REPAIR RATE

A more general approach is to take the repair time into consideration. At any time t  $\epsilon[0,T]$ , the computer is either at a functioning state or at a repair stage. Let us use X(t) to describe the state of the computer at time t. We define X(t) as

follows:

$$X(t) = \begin{cases} 1 & \text{if the machine is operating} \\ \\ 0 & \text{if the machine is being repaird.} \end{cases}$$

(5) 
$$\frac{dp_{1}(t)}{dt} = -\lambda(t)p_{1}(t) + \mu p_{0}(t)$$

(6) 
$$\frac{dp_0(t)}{dt} = -\mu p_0(t) + \lambda(t)p_1(t) ,$$

with the initial condition  $p_1(0) = 1$ . Since  $p_0(t) + p_1(t) = 1$ , it suffices to solve equation (5) only. Its solution is given by

(7) 
$$p_{1}(t) = e^{-\int_{0}^{t} [\lambda(z) + \mu] dz} \int_{0}^{t} e^{\int_{0}^{x} [\lambda(y) + \mu] dy} \mu dx + \int_{0}^{t} [\lambda(z) + \mu] dz$$

and

$$p_0(t) = 1 - p_1(t)$$

Hence EX(t) =  $p_1(t)$  and  $\sigma^2(X(t)) = p_1(t)[1-p_1(t)]$ . We have obtained the probability  $p_1(t)$  that the computer is operating at any time t.

Another statement can be made about the process X(t). Consider the stochastic integral  $\int_0^T X(t)dt$  which exists in quadratic mean. The integral  $\int_0^T X(t)dt$  represents the total amount of time that the process X(t) is in state 1 during [0,T]. Its expected value

$$E \int_0^T X(t)dt = \int_0^T p_1(t)dt$$

gives the expected length of the process (that the computer is operating) staying in state 1 during [0, T].

To obtain the probability distribution of the number of failures N(t) up to time t, we proceed as follows. First, let us re-draw Figure 1 of section 1 below

with 0 <  $T_1$  <  $\xi_1$  <  $T_2$  <  $\xi_2$  .... < T. According to our assumptions of the present section, the repair times  $V_i = \xi_i - T_i$  are independently distributed with the same exponential distribution

$$G(V) = 1 - e^{-\mu V}$$
  $V \ge 0 \quad \mu > 0$ 

and the conditional density of the i<sup>th</sup> failure time  $T_i$  given that the last recovery occurred at  $\xi_{i-1}$  is

(8) 
$$f(t_{1}/\xi_{1-1}) = \lambda(t_{1}) \exp \left\{-\int_{\xi_{1-1}}^{t_{1}} \lambda(z) dz\right\}, \ \lambda(z) > 0$$

$$t_{1} \ge 0$$

The joint probability density of the pair  $(T_i, \xi_i)$ , given  $\xi_{i-1}$ , is

(9) 
$$\lambda(t_i) \exp \left\{-\int_{\xi_{i-1}}^{t_i} \int_{\mu e}^{-\mu(\xi_i - t_i)} for \, \xi_i \geq t_i \geq \xi_{i-1} \right\}$$

= 
$$g(t_i, \xi_i/\xi_{i-1})$$
.

By simple probabilistic arguments, we obtain the joint probability distribution of the random variables  $T_1$ ,  $\xi_1$ ,  $T_2$ , ...,  $T_{N(t)}$  and N(t).

(10) 
$$\frac{N(t)}{\prod_{i=1}^{t}} \lambda(t_i) \exp \left\{-\int_{\xi_{i-1}}^{t_i} \lambda(z) dz\right\}_{\mu} \exp \left\{-(\xi_i - t_i)_{\mu}\right\} \cdot \exp\left\{-\int_{t_{N(t)}}^{t} \lambda(z) dz\right\}$$

with 0 <  $t_1$  <  $\xi_1$  <  $t_2$  <  $\xi_2$  ..... <  $t_{N(t)}$  <  $t_1$  Integrating equation (10) over the region 0 <  $t_1$  <  $\xi_1$  < ... <  $t_{N(t)}$  <  $t_1$  we get the probability distribution of N(t),  $P_n(t)$ . For the estimation of the parameters  $\mu$ , the problem is simple. The maximum likelihood estimate  $\hat{\mu}$  of  $\mu$  is given by

$$\hat{\mu} = \frac{N(t)}{N(t)}$$

$$\sum_{i=1}^{N(t)} (\xi_i - T_i)$$

To estimate  $\lambda(t)$  we need an explicit form for the function  $\lambda(t)$ .

### 5. RELIABILITY OF A REDUNDANT SYSTEM OF N IDENTICAL COMPUTERS

Suppose there is a system of N identical computers, each of which operates independently and performs the same type of calculations. The machines are subject to failures and repairs, with the failure rate  $\lambda(t)$  and repair rate  $\mu$  as defined in section 4. The system is said to be in state k, at time t, for  $0 \le k \le N$ , if there are k computers operating at time t while the remaining N-k machines are being replaced. We have considered the special case when N=1 in section 4, where the system is described by the stochastic process X(t). Now, for N 1, we introduce X\_i(t), i=1,...,N, with each X\_i(t) having the same probabilistic interpretation as that of X(t). The process X\_i(t) assumes value 1 or 0 at time t depending on whether the i<sup>th</sup> computer is operating or not.

Let us define  $S_N(t) = \sum_{i=1}^N X_i(t)$  which represents the number of machines operating at time t. Since for any fixed t,  $X_i(t)$  is a Bernoulli random variable with probability of being operable equal to  $p_1(t)$  (see equation (7)) and  $X_i(t)$ , for i=1,...,N are independently distributed, the sum  $S_N(t)$  follows a binomial distribution  $B(N,p_1(t))$ . The probability that k machines are operable at time t is given by

$$P[S_N(t)=k] = {N \choose k} p_1^k(t) [1-p_1(t)]^{N-k}$$
 for k=0,...,N .

If a minimum number of  $k^*$  machines is required to guarantee a satisfactory performance of the system, then, with a given N and  $p_1(t)$ , the reliability of the system at time t is expressed by the following probability.

$$P[S_{N}(t) \ge k*] = \sum_{i=k*}^{N} {N \choose i} p_{1}^{i}(t) [1-p_{1}(t)]^{N-i} = \alpha$$

On the other hand, given  $p_1(t)$  and  $\alpha$ , we are able to determine the number of machines N\* needed to achieve a predetermined reliability  $\alpha$ . We find the largest N\* such that

$$\inf_{0 \leq t \leq T} \ \text{P[S}_{N}(t) \geq k*] = \inf_{0 \leq t \leq T} \sum_{i=k*}^{N*} \binom{N*}{i} \ \text{p[t)[1-p[t)]}^{N*-i} \geq \alpha$$

### 6. CONCLUSIONS

The present paper derives a stochastic model useful in analoging the reliability of computer operations. At present, we used Bellcomm's UNIVAC 1108 hardware failure data as an example to study the reliability problem. It is found by the numerical evidence that the usual assumption of constant failure rate may not be applicable in setting up a reliability model. Thus a time-dependent failure rate  $\lambda(t)$  is introduced in building the model. It is also highly suspected that computer failure rate  $\lambda(t)$  may vary with its usage pattern. Accordingly, we would expect a different failure rate  $\lambda(t)$  in other computer systems. The model derived in this paper assumes no functional form for  $\lambda(t)$ , except the continuity and positivity conditions, this makes the model feasible for general applications. The study is being extended and will include analyses of MSC computer system data.

#### 7. ACKNOWLEDGMENT

I wish to thank Mr. J. Byrd and Mr. E. Hillyard for providing the Bellcomm UNIVAC 1108 hardware failure data.

1033-GLY-jr

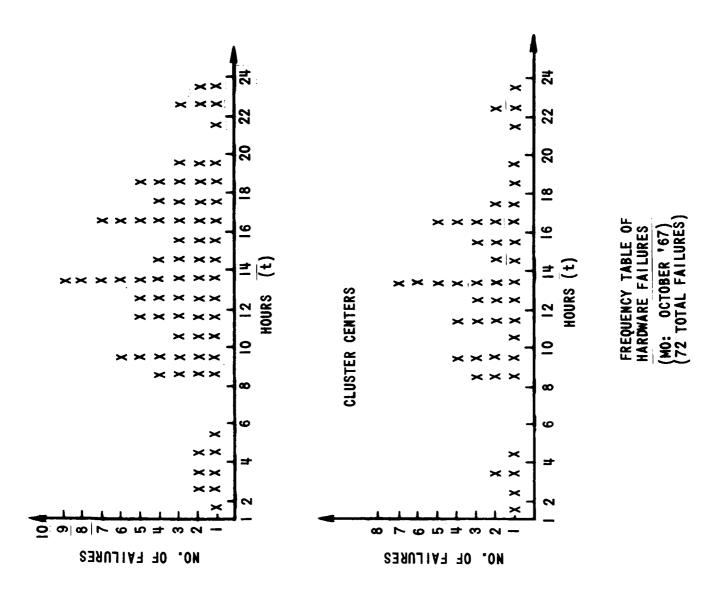
Gran J. Yang

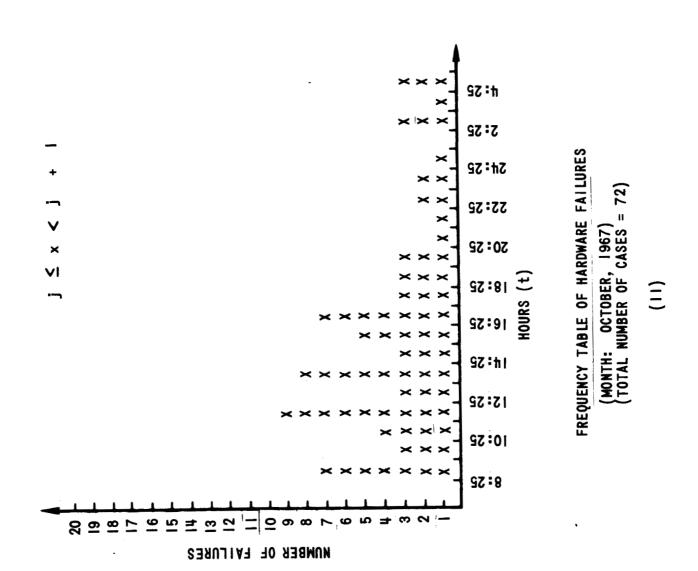
Attachments Reference Tables I - XI BELLCOMM, INC.

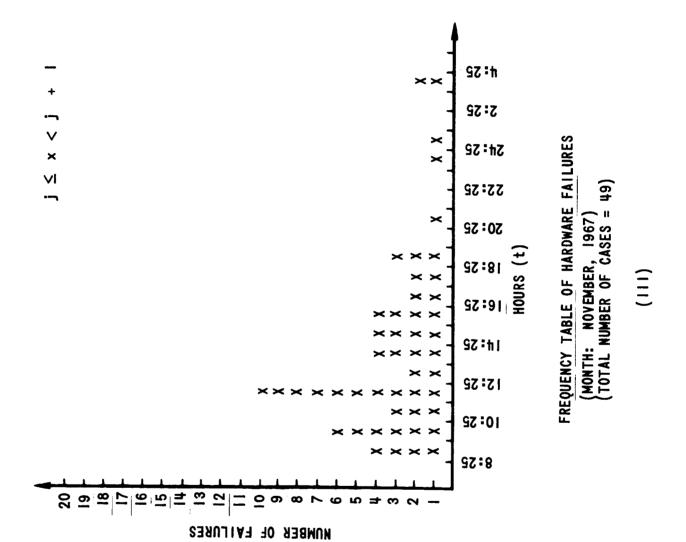
## REFERENCE

1. Pearson, E. S. and Hartley, H. O., <u>Biometrika Tables for Statisticians</u>, Vol. 1, Cambridge University Press, 1966, 3rd Edition.

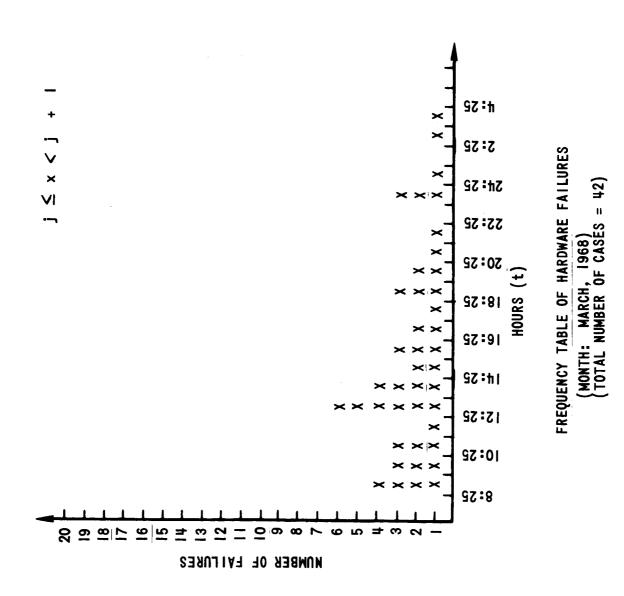


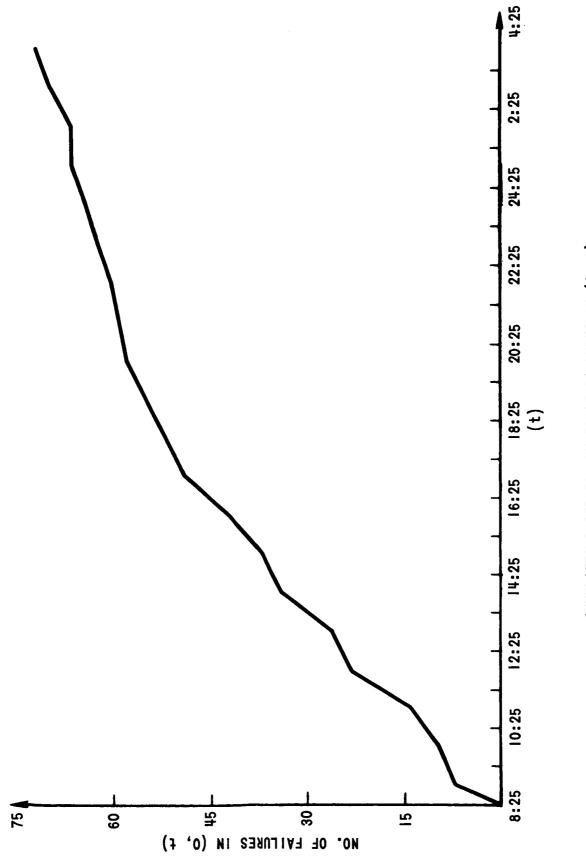






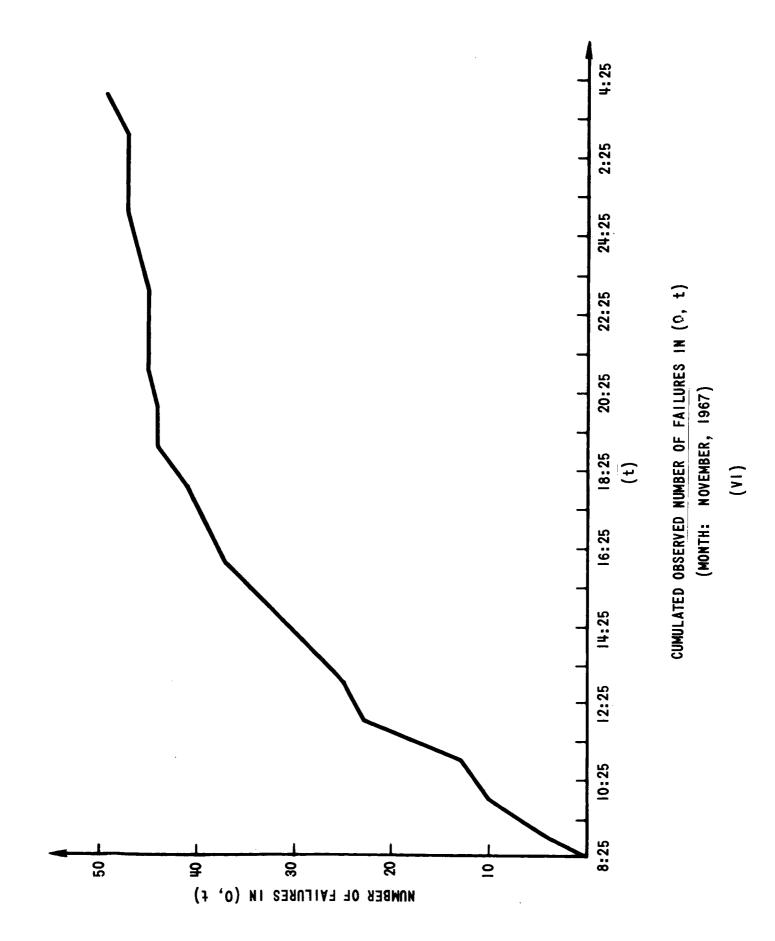


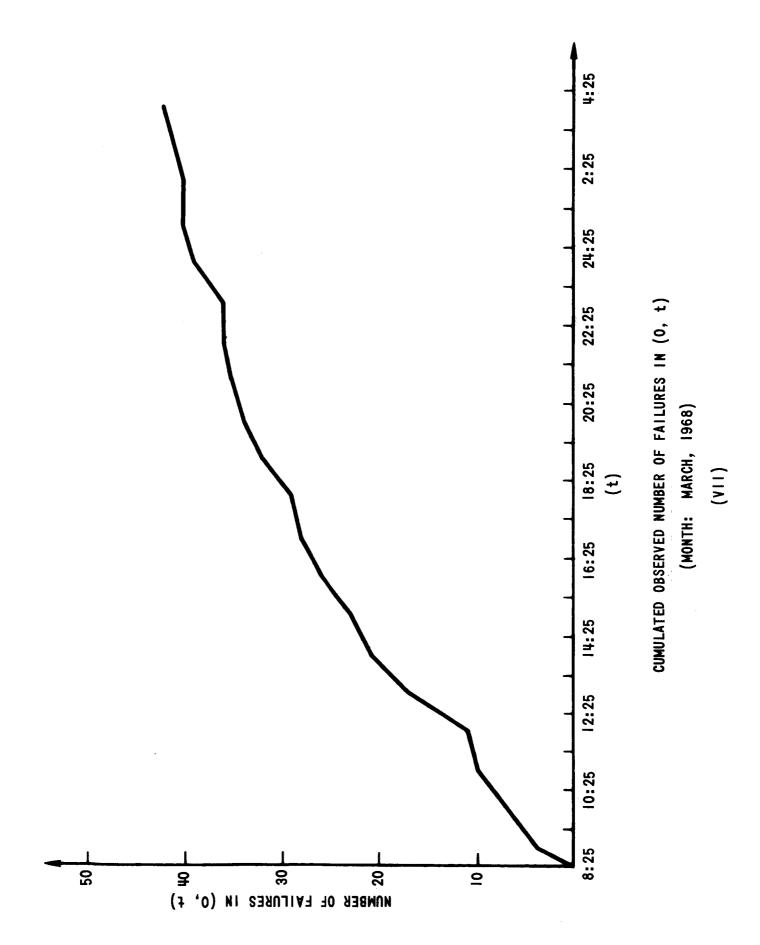


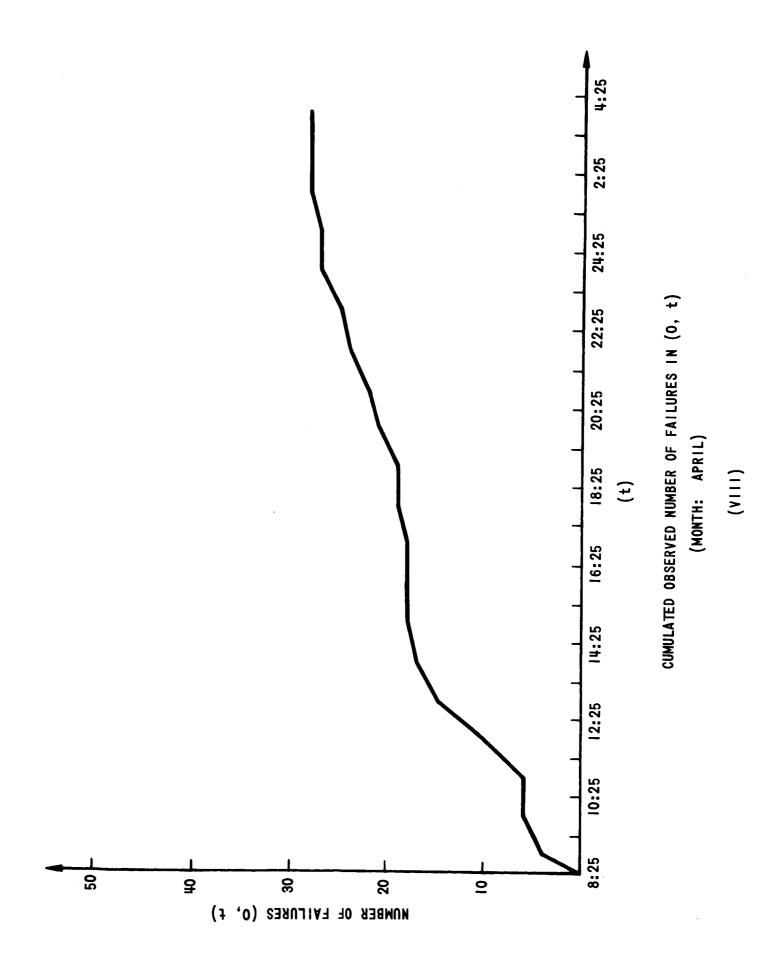


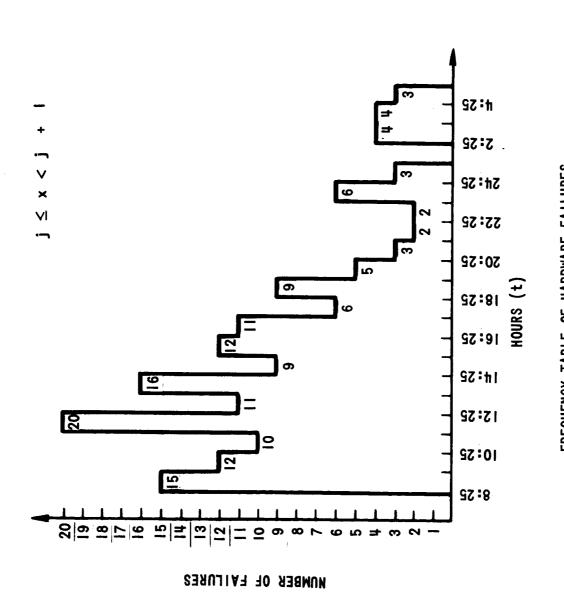
CUMULATED OBSERVED NUMBER OF FAILURES IN (0, t) (MONTH: OCTOBER)

5

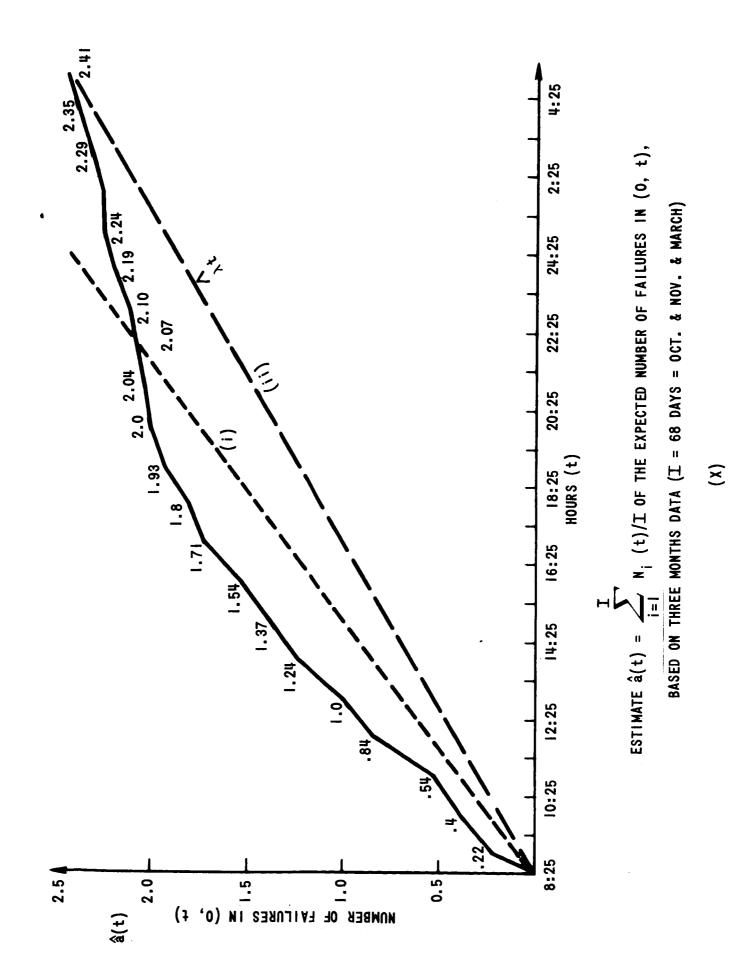








(MONTH: OCT. (26 DAYS) + NOV. (20 DAYS) + MARCH (22 DAYS) = 68 DAYS) (TOTAL NUMBER OF CASES = OCT. (72) + NOV. (49) + MARCH (42) = 163) FREQUENCY TABLE OF HARDWARE FAILURES



	<b>l</b> 1	ì			1	1		
DIFFERENCE	. 1213	.0227	. 0578	.0515	. 0242	1800.	.0021	
$\begin{array}{ccc} \lambda t &= 1 \\ (\lambda t)^n &= -\lambda t \\ && & \\ && & \\ && & \\ && & \\ && & \\$	. 3679	.3679	. 1839	.0613	.0153	. 003	. 0005	
$a(t) = 1.4$ $a^{n}(t) e^{-a(t)}$ $n!$	. 2466	.3452	.2417	.1128	.0395	1110.	. 0026	
P [N(t)=n] =	0 = u	1	2	3	ħ	5	9	